

How to evaluate an Early Warning System ?

*Towards a Unified Statistical Framework for Assessing Financial Crises
Forecasting Methods*

Bertrand Candelon* Elena-Ivona Dumitrescu† Christophe Hurlin‡§

January 2010

Abstract

This paper proposes a new statistical framework originating from the traditional credit-scoring literature, to evaluate currency crises Early Warning Systems (EWS). Based on an assessment of the predictive power of panel logit and Markov frameworks, the panel logit model is outperforming the Markov switching specifications. Furthermore, the introduction of forward-looking variables clearly improves the forecasting properties of the EWS. This improvement confirms the adequacy of the second generation crisis models in explaining the occurrence of crises.

Key words: currency crisis, Early Warning System, credit-scoring.

J.E.L Classification : C33, F37.

*b.candelon@algec.unimaas.nl, Maastricht University, School of Business and Economics, Department of Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands, Tel +31 433 88 34 42.

†elena.dumitrescu1@etu.univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), Rue de Blois, BP 6739 - 45067 ORLEANS CEDEX, France, +33 635 15 26 23.

‡christophe.hurlin@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), Rue de Blois, BP 6739 - 45067 ORLEANS CEDEX, France, +33 238 49 40 47.

§The authors thank the participants at the 3rd annual meeting of the MIFN (Luxemburg, October 2009), the 6^{eme} Journée d'économétrie "Développements récents de l'économétrie appliquée à la finance" (Paris, November 2009) and the 3rd FIW Research Conference "International Economics" (Vienna, December 2009) for their comments. The usual disclaimers apply.

1 Introduction

The recent sub-prime crisis has renewed interest in financial crises. In a prophetic paper, Bordo et al., (2001) distinguish four types of financial crises (*i.e.*, banking, currency, twin and all crises), and show that financial crises, in particular currency crises, would become more frequent after the collapse of the Bretton-Woods system in emerging as well as developed countries. This stylized fact should have stimulated economists to improve the quality of Early Warning Systems (EWS) set up to identify an impending currency crisis before it occurs. Such an alarm system constitutes the only tool available for authorities to implement optimal policies to prevent or at least attenuate the impact of a crisis.

The first EWS was proposed by Kaminsky, Lizondo and Reinhart, (1998) (KLR, hereafter) using a signaling approach. They use a large database of 15 indicator variables covering the external position, the financial sector, the real sector, the institutional structure and the fiscal policy of a particular country. An indicator (i) signals a crisis when it exceeds a particular cut-off point (C_i). The estimation of this threshold is at the core of such an analysis. KLR determine it to minimize the noise-to-signal ratio (NSR) such that the probability of the occurrence of a crisis is at a maximum after exceeding the cut-off point. The EWS for country j is then built as the weighting-sum of the individual indicators, the weights being given by the inverse of the NSR.

Berg and Patillo, (1999) (BP hereafter) compare this signaling method to a panel probit model. Their in-sample forecasts outperform KLR forecasts, when considering the NSR and measures similar to the mean square error, such as the quadratic probability score (QPS) and the log probability score (LPS). This analysis thus paved the way for a huge number of empirical studies (Kumar et al., 2003, Fuertes and Kalotychou, 2007, Berg et al., 2008). Nevertheless, in BP and all other studies, thresholds beyond which crises are detected are exogenously fixed, usually at 25% and 50%, and no formal statistical framework has been proposed to test for the improvement of the NtoRS. Hence, rather simplistically, a lower value is enough to conclude about the superiority of a model.

Moreover, these studies do not exploit the fact that turmoils refer to specific states structurally different from that observed during tranquil periods. Hence, Bussiere and Fratzscher, (2006) propose a multinomial logit EWS¹, whereas other studies use Markov-Switching models; see Abiad, (2003), Martinez-Peria, (2002) and Fratzcher, (2003). The evaluation of these models is similar to the one proposed in BP, as it relies on the NSR. The determination of the optimal cut-off point is not explicitly tackled, even if Harding and Pagan, (2006) show that it is endogenously determined. Nevertheless, in contrast to BP, there is no reason why the optimal cut-off point should lead to a minimum NSR.

1. They initially consider only two regimes (*i.e.*, turmoil and tranquil periods) but then integrate a third regime to characterize the post-crisis period, justifying it with recovery specificities.

Hence, even if these approaches seem to be different, they suffer from similar drawbacks in their evaluation. First, they all use the NSR measure as a so called *in fine* comparison criterion. As noticed by Bussiere and Fratzscher, (2006 p.957) this approach has a clear problem. If the cut-off point decreases, it will lead to a better detection of the coming crises (*i.e.*, the type 1 error will decrease), but at the same time, it will lead to an increase in the number of false alarms (*i.e.*, the type 2 error will increase). To summarize, a more efficient detection of future crises will occur at the cost of more frequent false alarms, which may have an efficiency cost in terms of economic policy. Second, no statistical inference is available to test for the forecasting superiority of an EWS when compared to another model. This represents an important issue in light of the previous remark. Namely, does a model exhibiting a lower type 1 but a higher type 2 error really outperforms the other models?

This paper aims to tackle both problems by developing a new statistical framework to evaluate EWS. Drawing on the traditional credit-scoring measure (Basel Committee on Banking Supervision, 2005 and Lambert and Lipkovich, 2008 *inter alii*), our proposed method goes beyond a simple analysis of the NSR, by determining the optimal threshold for each country, through reliance on the sensitivity-specificity plot and accuracy measures. Similarly, we adapt the most important credit-scoring criteria (including AUC, Kuiper Score, Pietra Index, and Bayesian Error Rate) to use them as evaluating criteria to gauge the forecasting performance of the models. Finally, inference for nested and non-nested hypothesis is developed to identify the optimal specification.

In the empirical section of this paper, this framework is implemented to compare the forecasting ability of a fixed-effects panel logit and a country-by-country Markov switching model for six Asian and six Latin-American countries. In addition, the role of forward-looking variables, which have a key role for the theoretical second generation models, according to Obstfeld (1996), is also investigated in both models. It appears that the logit model with market expectation variables outperforms the other logit or Markov specifications; based on optimally identified thresholds, this model forecasts correctly more than 67.9% of crises and 61% of calm periods in each of the twelve countries, and it is robust to certain changes in the construction of the dependent variable such as the specified currency crisis dating method or the choice of the forecasting horizon. In such a framework, the performance assessment criteria proposed and the comparison tests outperform the usual methods.

This paper is organized as follows: the panel logit as well as the Markov switching EWS models are presented in Section 2. Section 3 presents our new evaluation framework. The determination of an optimal cut-off as well as the evaluation criteria and comparison tests are developed. The database as well as the method used to date currency crises are scrutinized in Section 4. Empirical results are reported in Section 5 where the models are compared. Section 6 concludes.

2 The EWS Models: The Panel Logit and Markov Switching Models

In this section, the two most well-known EWS models (*i.e.*, panel logit and Markov switching models) are presented. They are later used in the empirical section of this paper to apply our new evaluation framework.

2.1 The Panel Logit Model

Following the seminal paper of BP, the fixed-effects panel logit model seems to constitute an adequate model for building an EWS. In a recent paper, Berg et al., (2008) advice against including all countries available in the panel, but instead suggest considering EWSs specific to homogenous clusters, which are determined in advance. The EWS for countries that cannot be included in any clusters rely on simple time series logit regressions.

Let $n = 1, \dots, N$ be the number of countries and $t_n = 1, \dots, T_n$ be the number of time periods considered for the n^{th} country. The dependent variable $y_{n,t_n,j}$ is a binary variable taking the value of 1 if at least one crisis occurs in the following j periods and 0 if no crisis occurs. j is the forecast horizon and will be removed from the notation for ease of simplicity.

Using a logistic cumulative distribution function, we obtain a conditional logit model for each cluster (Hsiao, 2003):

$$\Pr(y_{n,t_n} = 1) = \frac{\exp(\beta' x_{n,t_n} + f_n)}{1 + \exp(\beta' x_{n,t_n} + f_n)} \forall n \in \Omega_h, \quad (1)$$

where Ω_h is the h^{th} cluster, $h \in \{1, \dots, H\}$, and $\dim(\Omega_h) = N_h$, so that $\sum_{h=1}^H N_h = N$, and $\dim(\Omega_h)$ is the number of countries in the h^{th} cluster. f_n represents the fixed effects (*i.e.*, the constant term specific to each country).²

This panel specified as such exhibits several problems leading to the presence of serial correlation. First, Berg and Coke, (2000) show that considering a forecast horizon larger than 1 leads to autocorrelation in the crisis variable. Moreover, Harding and Pagan, (2006) show that including constructed binary variables in a model (as is the case for y_{n,t_n}) always leads to a serial correlation problem. A robust variance-covariance matrix is thus considered using the sandwich estimator (Williams, 2000).³

Nevertheless, the binary choice models have some major drawbacks. First of all, the dependent variable is a binary one, requiring an a priori dating of crisis periods, which

2. In the case of time-series models there are no individual effects and thus, we can estimate a standard unconditional bivariate logit. The estimation is straightforward.

3. This method is described in Appendix 1.

is always challenging. Second, the threshold used to distinguish the crisis from non-crisis periods are arbitrarily chosen. Third, a part of information is lost when a continuous variable is transformed into a qualitative one.

2.2 The Markov Switching Model

Markov switching models do not require a prior dating of crises and impose fewer distributional hypotheses.

Proposed first by Hamilton, (1988, 1995) so as to analyze the stance of business cycles, this model assumes regime-specific relationship between variables. The transition between states is endogenous and depends on fixed transition probabilities⁴ that are estimated via a modified Maximum Likelihood (ML) approach. The model used in the paper allows for changes in regime, in regressors, as well as in volatility (Abiad, 2003); it has the following form:

$$KLRm_t = \mu_t(S_t/\mathfrak{S}_t) + \beta(S_t/\mathfrak{S}_t)x_t(S_t/\mathfrak{S}_t) + \epsilon_t(S_t/\mathfrak{S}_t), \quad (2)$$

where $KLRm_t$ is a market pressure index vector, x_t represents the matrix of economic variables, μ is the intercept, ϵ_t is *i.i.d.*($0, \sigma_{\mathfrak{S}_t}^2$) and \mathfrak{S}_t represents the information set available at time t . S_t is a latent variable representing the regime, which follows a first order two state Markov chain $\{S_t\}_{t=1}^T$. $S_t = 0$ if there is a crisis, and $S_t = 1$ if not. For the ex-post identification of the two regimes, Abiad, (2003) considers that the crisis (tranquil) regime is that with higher (lower) volatility.

The constant transition probability matrix between these regimes from time $t - 1$ to t is:

$$P = \begin{pmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{pmatrix} = \begin{pmatrix} P_{00} & 1 - P_{11} \\ 1 - P_{00} & P_{11} \end{pmatrix}, \quad (3)$$

The model is estimated by Maximum Likelihood, as described by Hamilton, (1995). The initial values of the parameters are obtained from Ordinary Least Squares (OLS) regressions. For each state, the conditional mean and difference to the mean are computed. Next, the normal probability density for each regime η_t is obtained. Given the initial values of the parameters μ_0 , β_0 and of the conditional probability for each regime $\xi_0 = 1/2$, we can iterate from $t = 1$ to T on the following equations:

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \circ \eta'_t}{1'(\hat{\xi}_{t|t-1} \circ \eta'_t)} \quad (4)$$

4. Time-varying transition probabilities have been proposed by Diebold, Weinbach and Lee (1994) but are not considered in the paper.

$$\hat{\xi}_{t|t-1} = P\hat{\xi}'_{t-1|t-1}, \quad (5)$$

where \circ denotes element-by-element multiplication, and 1 is the unit vector (*i.e.*, a column vector).

The first equation yields the filtered probabilities⁵ $\Pr(S_t = i|\mathfrak{S}_t)$ for each state, while the second equation shows the forecasted probabilities of being in such a state in the next period $\Pr(S_{t+1} = i|\mathfrak{S}_t)$. The conditional normal density η_t and the filtered probabilities allow for the computation of the conditional log likelihood of the observed data:

$$L(\theta) = \sum_{t=1}^T \log(1'(\hat{\xi}_{t|t-1} \circ \eta_t)). \quad (6)$$

Since our objective is to compare the results of the logit and Markov methods, we must obtain a series of j -month-ahead forecasts. More precisely, we estimate the probability of observing at least one crisis in the next j periods as follows (Arias and Erlandsson, 2005):⁶

$$\Pr(S_{t+1\dots t+j} = 0|\mathfrak{S}_t) = 1 - \Pr(S_{t+1\dots t+j} = 1|\mathfrak{S}_t) \quad (7)$$

$$= 1 - \{[P_{01}P_{11}^{(j-1)}\Pr(S_t = 0|\mathfrak{S}_t)] + [P_{11}^j\Pr(S_t = 1|\mathfrak{S}_t)]\}. \quad (8)$$

3 A New Framework for Evaluating EWS Models

The evaluation of EWS models is particularly important, since several models have been proposed. However, most of the studies comply with only a few performance assessment procedures based on an arbitrary choice of cut-offs and do not offer a framework allowing for statistical inference.

In this section, we propose a new evaluation framework that draws on the credit-scoring literature. In the first step, operating characteristics, *i.e.*, accuracy measures and misclassification errors, are presented and utilized to determine optimal cut-off points. In the second step, evaluation measures are developed to gauge the forecasting quality of the models. Finally, a battery of tests is presented to test for the best performing model.

3.1 Performance Assessment and Optimal Cut-off

This section presents the criteria used to compare the crises probabilities obtained from an EWS model with the actual occurrence of crises within a certain horizon. To be more precise,

5. We are often interested in forming an inference about the true regime at date t based on observations obtained up to a later date T , denoted by $\xi_{t|T}$. These are referred to as "smoothed" probabilities $\Pr(S_t = i|\mathfrak{S}_T)$, and they are given by Kim's algorithm (1994): $\xi_{t|T} = \xi_{t+1|T}\xi_{t|t}P/\xi_{t+1|t}$.

6. See Krolzig, (2000) for a description of other estimation methods.

the estimated EWS models output crisis probabilities, and to shift from these probabilities to crisis forecasts, we must define an optimal threshold (or cut-off) that discriminates between predicted crisis periods and predicted calm periods. If the probability of a crisis is greater than the cut-off, the model issues a signal of a forthcoming crisis. Moreover, the lower the threshold is, the more signals the model will send (*i.e.*, the type 1 error will decrease), but at the same time, the number of wrong signals rises (*i.e.*, type 2 error will increase). Conversely, using a higher threshold level reduces the number of wrong signals, but the number of missing crisis signals increases. Therefore, we can define an indicator variable of predicted occurrence of crises for the next j periods as following:

$$\hat{I}_t = \begin{cases} 1, & \text{if } P_t > C \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

where C represents a fixed cut-off.

We address this trade-off by following the methodology used in the credit-scoring literature (Basel Committee on Banking Supervision, 2005). In fact, several evaluation criteria of predictive ability are available in the microeconomic risk assessment literature, but we concentrate on the most important ones, namely, Kuiper's score, Quadratic Probability Score, Log Probability Score, the area under the ROC curve (AUC), Pietra Index, and Bayesian Error.

First, there will be presented a few basic concepts and the two methods we use to determine the optimum threshold for each country. For comparison purposes, we also report the KLR cut-off introduced by Kaminsky et al., (1998). Then, we introduce the performance assessment criteria.

3.1.1 Optimal Cut-off Identification

For a given value of the cut-off C , where $C \in [0, 1]$, Table 1 reports the link between the observed y_t and predicted \hat{I}_t conditions in the following matrix:

TABLE 1 – True versus predicted occurrence of crises

		True value		
		Crisis	No crisis	Total
Predicted result	Crisis	True Positive (TP(C))	False Positive (FP(C))	All predicted crisis
	No crisis	False Negative (FN(C))	True Negative (TN(C))	All predicted non-crisis
	Total	All true crisis (N_D)	All non-crises (N_{ND})	T (sample size)

Note : For a given cut-off we count the number of crises correctly identified and the number of non-crises well predicted by the model (the first diagonal elements of the matrix) and the misidentified crisis and respectively non-crisis (the off diagonal elements). The number of real crises/non-crises periods lies on the 'total' row, while the number of predicted crises/non-crises lies on the 'total' column.

BP consider the NSR to be the only objective criteria; it is defined as:

$$NSR(C) = \frac{FP(C)}{TP(C)}. \quad (10)$$

The optimal threshold C^* is obtained as the value that minimizes $NSR(C)$; see BP and KLR.

It is nevertheless possible to go deeper, defining first the hit rate *i.e.*, *sensitivity* as

$$HR(C) = \frac{TP(C)}{N_D}, \quad (11)$$

where $TP(C)$ is the number of crises predicted correctly (*i.e.*, hits) using a cut-off equal to C , and N_D is the total number of crises in the sample. At the same time, the false alarm rate *i.e.*, $1 - specificity$ is defined as

$$FAR(C) = \frac{FP(C)}{N_{ND}}, \quad (12)$$

where $FP(C)$ is the number of false alarms (or false signals), and N_{ND} is the total number of non-crisis periods. The optimal cut-off rate is determined so as to maximize simultaneously and conditionally *sensitivity* and *specificity* (see Figure 1).

insert Figure 1

Several additional measures of accuracy based on operating characteristic have been implemented (Lambert and Lipkovich, 2008), including the Youden Index (J), Total Accuracy (TA), and the Matthews Correlation Coefficient (MCC). In addition, the total and weighted misclassification errors are presented (TME and WME). Similarly to *sensitivity* and *specificity*, these accuracy and error measures can be used to define the optimal cut-off value. More precisely, the TA measure is defined as the proportion of cases correctly predicted:

$$TA = \frac{TP(C) + TN(C)}{T}, \quad (13)$$

the Youden Index is defined by:

$$J = HR(C) - FAR(C), \quad (14)$$

and finally, the Matthews Correlation Coefficient is defined as:

$$MCC = \frac{TP(C) * TN(C) - FP(C) * FN(C)}{\sqrt{((TP(C) + FN(C)) * (TP(C) + FP(C)) * (TN(C) + FP(C)) * (TN(C) + FN(C)))}}. \quad (15)$$

At the same time, the misclassification errors are calculated as follows:

$$TME = FN(C) * L_{fn} + FP(C) * L_{fp}, \quad (16)$$

$$WME = \frac{FN(C) + W_{fn}}{T} + \frac{FP(C) + W_{fp}}{T}, \quad (17)$$

where L_{fn} are the losses associated with crisis periods categorized as tranquil ($L_{fn} = 1/N_D$), L_{fp} are the losses associated with calm periods identified as crises ($L_{fp} = 1/N_{ND}$), and W_{fn} and W_{fp} are weights based on the relative losses, namely, $W_{fn} = L_{fn}/(L_{fn} + L_{fp})$, and $W_{fp} = L_{fp}/(L_{fn} + L_{fp})$.

As a result, it is possible to determine for each country an optimal cut-off for which the accuracy measures are maximized and error measures are minimized.

3.1.2 Evaluation Criteria

Furthermore, we provide more details on the performance assessment criteria previously mentioned at the beginning of this subsection. To start with, we can formally introduce Kuiper's score (Granger and Pesaran, (2000)) as the difference between the proportion of crises correctly predicted HR and the proportion of tranquil periods incorrectly forecasted FAR :

$$KS = HR - FAR. \quad (18)$$

The model above generates more hits (that is, correctly identified crises) than false alarms if the value of Kuiper's score is positive.

The Quadratic Probability Score (QPS) is a mean square error measure that compares the predicted probabilities of the two states (*i.e.*, crisis and non-crisis) with the real crisis indicator. It is defined as:

$$QPS = \frac{1}{T} \sum_{t=1}^T 2(P_t - y_t)^2, \quad (19)$$

where P_t represents the estimated probability of crisis at time t and y_t is the realization of the crisis event at time t . QPS takes values from 0 to 2, with 0 indicating perfect accuracy.

In contrast, the Log Probability Score (LPS) loss function penalizes large errors more heavily than QPS:

$$LPS = -\frac{1}{T} \sum_{t=1}^T ((1 - y_t)\ln(1 - P_t) + y_t\ln(P_t)). \quad (20)$$

This score ranges from 0 to ∞ , with LPS=0 being perfect accuracy.

Before addressing the area under the ROC Curve (AUC) performance assessment criteria, it is necessary to define the Receiving Operating Characteristic (ROC) curve. The ROC curve is a visual tool whose concavity is equivalent to the conditional probabilities of a crisis being a decreasing function of the underlying scores. Its objective is to minimize misclassification and maximize the overall hit rate of the model. More exactly, the ROC curve is a graphic representation of the *sensitivity* and $1 - \textit{specificity}$ relative to the cut-off C , where $C \in [0, 1]$ (see Figure 2). For a perfect model, the curve passes through the point $(0,1)$, indicating that it correctly recognizes all crisis and calm periods.

insert Figure 2

The area under the ROC curve (AUC) measure, which ranges from 0 to 1, provides a measure of the model's overall ability to discriminate between the cases correctly predicted and the false alarms. The larger the AUC, the better the model. This statistic can be calculated as $A = \int_0^1 HR(FAR)d(FAR)$, and its value corresponds to the Wilcoxon-Mann-Whitney ranking statistic. In other words, the AUC estimates the probability that a randomly chosen crisis observation is ranked higher than a non-crisis observation. Thus, a perfect ranking means that all crisis observations are ranked higher than the non-crisis observations, and consequently, it implies an AUC equal to 1. In contrast, the expected value of the AUC statistic for a random ranking is 0.5.

The Pietra Index is another ROC measure which quantifies half of the maximal distance between the ROC and the diagonal of the unit square. Like the AUC, the Pietra Index does not depend on the crisis probability of the sample. Geometrically, in the case of a concave ROC curve, we obtain the following representation of the statistic:

$$Pietra\ Index = \frac{\sqrt{2}}{4} \max_C |HR(C) - FAR(C)|. \quad (21)$$

Last but not least, the Bayesian Error rate (or classification error) criteria returns the minimum probability of error for a binary case model. It can be estimated parametrically or non-parametrically, but in the case of a concave ROC curve the error rate can be expressed as:

$$Error\ rate = \min_C (P_D(1 - HR(C)) + (1 - P_D)FAR(C)), \quad (22)$$

where P_D represents the rate of crisis in the sample (*i.e.*, $P_D = N_D/T$).

3.2 Comparison Tests

So far, we have developed methods to obtain an optimal cut-off as well as to evaluate forecasts. While most of the other studies would stop at this stage, and select the model providing the best forecast evaluation, we think it is necessary to develop a suitable statistical framework to test for the forecasting equivalence between two models with respect to the evaluation measure.

Accordingly, we implement three such tests, namely, the Diebold-Mariano (DM) statistic for non-nested models (Diebold and Mariano, 1995, Clark and West's, 2007 'MSPE-adj' statistic for nested models, and a test of comparison of correlated ROC curves DeLong et al., 1988.

The first test was proposed by Diebold and Mariano (1995); it can be built on any criteria, including MSFE, MAE, or even loss functions.

Let us consider two forecasts at a horizon k , $\hat{y}_{1,t+k}$, and $\hat{y}_{2,t+k}$, $t \in \{1, \dots, T\}$ of the time series y_t , obtained from two concurrent models. Now, let the forecast errors for the two models be $\{\widehat{e}_{1,t}\}_{t=1}^T$, and $\{\widehat{e}_{2,t}\}_{t=1}^T$. The loss function associated with a forecast e_t is denoted by $g(\widehat{e}_t)$. Now, we can formulate the null hypothesis of equal forecast accuracy as follows:

$$H_0 : E[g(\widehat{e}_{1,t})] = E[g(\widehat{e}_{2,t})], \quad (23)$$

or

$$H_0 : E(d_t) = 0, \quad (24)$$

where d_t represents the loss differential: $d_t = g(e_{1,t}) - g(e_{2,t})$.

Let us define by \bar{d} the loss differential mean $\bar{d} = (1/T) \sum_{t=1}^T d_t$. Under the null hypothesis, \bar{d} follows a normal distribution with a variance equal to $\sigma_{\bar{d},0}^2/T$, where $\sigma_{\bar{d},0}^2$ is the long term variance of the loss differential.

Thus, the DM test statistic based on the loss differential mean takes the following form:

$$DM = \frac{\bar{d}}{\sqrt{\text{var}(\bar{d})/T}} = \frac{\bar{d}}{\sigma_{\bar{d},0}\sqrt{T}} \xrightarrow[T \rightarrow \infty]{d} N(0, 1). \quad (25)$$

The long term variance can be obtained by using a kernel estimator which corresponds to a weighted sum of future and past autocovariances of the loss differential d_t . In the case of a uniform kernel as chosen by Diebold and Mariano, (1995), the long term variance estimator has the following form:

$$\sigma_{\bar{d},0}^2 = \sum_{t=-(k-1)}^{k-1} \hat{\gamma}_d(j) = \hat{\gamma}_d(0) + 2 \sum_{t=1}^{k-1} \hat{\gamma}_d(t), \quad (26)$$

where $\hat{\gamma}_d(t) = (1/T) \sum_{i=t+1}^T (d_i - \bar{d})(d_{i-t} - \bar{d})$ is the empirical autocovariance of order t of d_t .

The authors highlight that the auto-covariances can be at most of order $k - 1$, where k is the forecast horizon. Therefore, in the case of one-step-ahead forecasts, the variance estimator is given by the empirical variance of the loss differential $\hat{\gamma}_0(t)$. In addition, since there is no particular loss function that can be used in the case of dichotomous dependent variables, the MSFE measure is used as the comparison criterion of the models. As a result, the loss functions become $g(\hat{e}_{1,t}) = \hat{e}_{1,t}^2$ and $g(\hat{e}_{2,t}) = \hat{e}_{2,t}^2$. The rest of the equations can be obtained by a simple substitution.

Even if this test is appealing with respect to our analysis, it cannot be implemented when models are nested. In this case, the distribution under the null hypothesis can no longer be established. An appropriate alternative test has been suggested by Clark and McCracken, (2001) and Clark and West, (2007).

Let us consider that model 1 is the parsimonious model and model 2 is a larger one, that reduces to model 1 if some of its parameters are set to 0. We remind the reader that the sample size is T . In addition, the k step ahead forecasts of the two models are denoted by $\hat{y}_{1,t+k}$ and $\hat{y}_{2,t+k}$.

The null hypothesis is the equality of MSPEs across the models, while the alternative is that the unrestricted model (*i.e.*, model 2) has a smaller MSPE than the restricted model 1; *i.e.*, model 2 outperforms model one.

Consequently, we can compute Clark and West's, (2007) MSPE-adj. statistic as follows:

$$\text{MSPE-adj.} = \frac{\sqrt{T}\bar{f}}{\sqrt{\hat{V}}}, \quad (27)$$

where $\hat{f}_{t+k} = (y_{t+k} - \hat{y}_{1,t+k})^2 - [(y_{t+k} - \hat{y}_{2,t+k})^2 - (\hat{y}_{2,t+k} - \hat{y}_{1,t+k})^2]$, \bar{f}_{t+k} is the sample average of \hat{f}_{t+k} and \hat{V} is the sample variance of $\hat{f}_{t+k} - \bar{f}$. This one-sided test uses critical values from the standard normal distribution.

The final test is based on DeLong et al., (1988); This is a study that presents a nonparametric analysis of areas under correlated ROC curves by using the theory on generalized U-statistics to generate an estimated covariance matrix. Let us denote by \widetilde{AUC}_1 and \widetilde{AUC}_2 the areas under the ROC curves associated with two EWS models. The null hypothesis is the equality of areas under the ROC curve *i.e.*, $\widetilde{AUC}_1 = \widetilde{AUC}_2$; in other words, neither of the models performs better than the other.

Then, we can introduce the test statistic as having the following form ⁷:

$$\frac{(\widetilde{AUC1} - \widetilde{AUC2})^2}{Var(\widetilde{AUC1} - \widetilde{AUC2})} \xrightarrow{T \rightarrow \infty} \chi^2(1), \quad (28)$$

where the variance term is based on the elements of the variance-covariance matrix S of the $\widetilde{AUC} = (\widetilde{AUC1}, \widetilde{AUC2})$ vector: $Var(\widetilde{AUC1} - \widetilde{AUC2}) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$. In addition, this covariance matrix S , which is a key component of the test, can be expressed as:

$$S = \frac{1}{N_D} S_{10} + \frac{1}{N_{ND}} S_{01}. \quad (29)$$

Note that S_{10} and S_{01} are defined on the basis of Hoeffding's, (1948) theory for generalized U-statistics. This statistic has a χ^2 distribution with one degree of freedom. In addition, we remind the reader that in small samples, the power of the test is likely to be small.

4 Data

Before before detailing the model comparison, it is important to look at the data as well as some specification problems, such as the way in which the binary variable is built and data poolability).

4.1 Dataset

The dataset is similar to the one used in Berg et al., (2008): It covers 12 countries ⁸ spans the period from January 1985 to January 2005 and is extracted from the IMF-IFS database as well as the national banks of the countries under analysis via Datastream. The series are provided on a monthly frequency, denoted in US dollars and adjusted for seasonality.

Several explanatory variables from three economic sectors are considered (Lestano et al., 2003):

1. External sector: the one-year growth rate of international reserves, the one-year growth rate of imports, the one-year growth rate of exports, the ratio of M2 to foreign reserves, and the one-year growth rate of M2 to foreign reserves.

7. This test can easily be extended to compare h different models: $(\widetilde{AUC} - AUC)L'[LSL']^{-1}L(\widetilde{AUC} - AUC) \xrightarrow{T \rightarrow \infty} \chi^2(df)$, where $\widetilde{AUC} = (\widetilde{AUC}_1, \widetilde{AUC}_2, \dots, \widetilde{AUC}_h)$, L is the coefficients vector (row) and S is the covariance matrix of the \widetilde{AUC} vector. The statistic follows a χ^2 distribution, where the number of degrees of freedom df is given by the rank of the matrix LSL' .

8. Argentina, Brazil, Mexico, Peru, Uruguay, Venezuela, Indonesia, South Korea, Malaysia, Philippines, Taiwan and Thailand

2. Financial sector: the one-year growth rate of M2 multiplier, the one-year growth rate of domestic credit over GDP, the one-year growth rate of real bank deposits, the real interest rate, the lending rate over deposit rate, and the real interest rate differential.

3. Domestic real and public sector: the one-year growth rate of industrial production.

We also add variables which incorporate a forward-looking dimension. Such a choice allows to gauge the possible self-fulfilling origins of currency crises (Obstfeld, 1984). Thus, the term spread (*i.e.*, long term government bonds minus the short term money rate) is introduced, as it indicates the feeling of the market with respect to future inflation and output growth (Estrella and Hardouvelis, 1991). In case of missing data, proxies like 'money ten days,' 'interbank one year rate,' and 'money 364 days' were used. To be more precise, the yield spread is actually a forward interest rate that can be decomposed into an expected real interest rate and an expected inflation component (Estrella and Mishkin, 1996). Thus, on the one hand, it is associated with expectations of future monetary policy and accordingly of future real growth, while on the other hand, expected inflation is related to changes in investors' expectations. Most importantly, since the yield spread seems to outperform other variables at long term forecasting horizons that are relevant from an investor's point of view, this variable is more forward-looking than other leading indicators (Estrella and Hardouvelis, 1991, Estrella and Trubin, 2006). Consequently, the use of the yield spread as a forecasting tool is even more compelling since it can issue a crisis signal well in advance of other indicators. Similarly, market confidence can be perceived through the stock market prices, since they are determined by expectations about future dividend streams and thus related to the future state of the economy.

As in Kumar, (2003), we reduce the impact of extreme values by using the formula: $f(x_t) = \text{sign}(x_t) * \ln(1 + |x_t|)$, so as to . Traditional first generation (Im, Pesaran, Shin, 1997 and Maddala and Wu, 1999) and second generation (Bai and Ng, 2000 and Pesaran, 2003) panel unit root tests are performed, leading to the rejection of the null hypothesis of stochastic trend except for the lending rate over deposit rate indicator. Hence, this series is substituted by its first difference.

Finally, we identify the most correlated leading indicators. The real interest rate is highly correlated with the real interest rate differential, while one-year growth of imports is strongly correlated with the one-year growth of industrial production as well as with the one-year growth of exports. Based on the minimization of the AIC and BIC information criteria of the panel data models, we identify the best two variables, namely, the real interest rate and the one-year growth of industrial production. The gaps through the series are replaced with mean values, but when the series revealed missing values at the beginning of the sample, such as "the one-year growth of terms of trade" or "yield spread", the corresponding observations are dropped from the analysis, leading to an unbalanced panel framework.

4.2 Dating Currency Crises

A crisis episode is generally detected if an index of speculative pressure exceeds a certain threshold. Many alternative indexes have been developed and used for identifying crises. But they are all non-parametric termination rules that take into consideration the size of the movements in a combination of a number of series. Lestano and Jacobs, (2004) compare several currency crisis dating methods, aiming to identify the one that recognizes most of the crises categorized by the IMF. They conclude that the KLR modified index, the Zhang original index (Zhang, 2001), and extreme values applied to the KLR modified index perform best.

Following their results, we identify crisis periods using the KLR modified pressure index (KLRm), which, unlike the KLR index, also includes interest rates:

$$\text{KLRm}_{n,t} = \frac{\Delta e_{n,t}}{e_{n,t}} - \frac{\sigma_e}{\sigma_r} \frac{\Delta r_{n,t}}{r_{n,t}} + \frac{\sigma_e}{\sigma_i} \Delta i_{n,t}, \quad (30)$$

where $e_{n,t}$ denotes the exchange rate (*i.e.*, units of country n 's currency per US dollar in period t), $r_{n,t}$ represents the foreign reserves of country n in period t , while $i_{n,t}$ is the interest rate in country n at time t . Meanwhile, the standard deviations σ_X are actually the standard deviations of the relative changes in the variables $\sigma_{(\Delta X_{n,t}/X_{n,t})}$, where X denotes each variable separately, including the exchange rate, foreign reserves, and the interest rate, with $\Delta X_{n,t} = X_{n,t} - X_{n,t-6}$.⁹

For both subsamples, the threshold equals two standard deviations above the mean¹⁰:

$$\text{Crisis}_{n,t} = \begin{cases} 1, & \text{if } \text{KLRm}_{n,t} > 2\sigma_{\text{KLRm}_{n,t}} + \mu_{\text{KLRm}_{n,t}} \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

From a macroeconomic point of view, it is more important to know if there will be a crisis in a certain horizon than in a certain month, because this time period allows the state to take steps to prevent the crisis. Consequently, we define for each country $C24_t$, which corresponds to y_t from our general framework and thus serves as the crisis dummy variable

9. Additionally, we take into account the existence of higher volatility in periods of high inflation, and consequently the sample is split into high and low inflation periods. The cut-off corresponds to a six month inflation rate higher than 50%.

10. In the case of KLR the threshold equals three standard deviations; however, in this case, Taiwan would never register any currency crises, which is historically not accurate. For example, Taiwan was not exempted from the Asian crisis in 1997.

taking the value of 1 if there will be a crisis in the following 24 months and 0 otherwise ¹¹:

$$C24_{n,t} = \begin{cases} 1, & \text{if } \sum_{j=1}^{24} Crisis_{n,t+j} > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (32)$$

It is important to note that in the case of Markov switching model, since the two regimes are intrinsically identified by the model, we use the KLR modified index as a continuous variable, rather than transform it into a binary one. Nevertheless, the identification of crisis dates concerns us from the viewpoint of forecast evaluation criteria and model comparison tests.

4.3 Optimal Country Clusters

Berg et al., (2008) have pointed out the importance of applying a panel-logit model only on optimal country clusters. To identify them, we use the Kapetanios sequential procedure (Kapetanios, 2003) much like Berg et al., (2008). Unsurprisingly, four optimal clusters are identified, namely, Argentina, Brazil, Mexico, Venezuela; Peru and Uruguay; Korea, Malaysia, Taiwan; Philippines and Thailand, and one non-poolable country, Indonesia. This existence of country clusters confirms the fact that in different countries there are different factors explaining currency crises. In fact, we expect to see clusters of countries smaller than the regional categories (*i.e.*, Latin America and South Asia), since they have experienced similar macroeconomic events (*i.e.*, high inflation rates, reforms, and stabilization plans) and are linked by strong trading relations.

In search of comparable results, we shall use the five ¹² previously identified clusters for all panel and time-series data models developed in this paper.

5 Empirical Results

As aforementioned, the aim of this paper is to evaluate which model best identifies crisis periods as crises and calm periods as calm periods. To do this, we use both a panel logit and a Markov switching framework. More exactly, we develop several specifications for each of the two approaches, including a logit model without market expectation variables, a logit model with market expectation variables, two Markov models without market expectation variables (*i.e.*, one with intercept and standard error switching coefficients and one including

11. The entire exercise is performed with C12 in order to test for the robustness of this approach in Appendix 2. The results show that outcomes are similar.

12. We consider the non-poolable country (that is, Indonesia) as an individual cluster. Therefore, five optimal clusters are identified.

an additional *international reserves* - hereafter *resg* switching coefficient), and three Markov models with market expectation variables (*i.e.*, one with intercept and standard error switching coefficients, a second one that includes an *international reserves* switching coefficient, and a third one that includes a *yield spread* - hereafter *spread* switching coefficient as opposed to the two previously mentioned models). Then, for each model we opt for the following 4-steps approach:

First, each model is estimated so as to obtain the country probability at time t of having a crisis in the following 24 months.

Second, the optimal cut-off is identified by applying our new backtesting method based on credit-scoring criteria.

Third, the predictive performance of each model is scrutinized.

Finally, comparison tests are implemented so as to find the model that overperforms the others.

5.1 Estimation, Optimal Cut-off and Performance Assessment

In this section, we analyze the estimation results of our models, and for a given model, we identify the optimal cut-off and compute the values of the performance assessment criteria for each of the 12 countries. Since we study numerous specifications, the results are only partially reproduced in the paper ; the entire results are available upon request.

insert Tables 2 and 3

Tables 2 and 3 show the Maximum Likelihood estimates of the logit model with market expectation variables for each of the five optimal clusters. The goodness of fit indicators reveal that the independent variables have important explanatory power for all clusters. Moreover, the signs of the coefficients tend to correspond to a priori expectations. Nonetheless, we encounter some variations from one cluster to another concerning both the significance and the sign of the parameters. Therefore, we find evidence of parameter heterogeneity across optimal clusters. To be more precise, only the growth of international reserves and the growth of M2 to reserves coefficients are always negative, indicating that a rise of one unit of their value implies a reduction in the crisis probability. Besides, the signs of the coefficients for variables like growth of domestic credit over GDP, growth of the stock market price index, or yield spread are relatively stable across the five clusters, as only one of the signs differs.

At the same time, we identify growth of real bank deposits, growth of M2 to reserves and yield spread as being the most significant factors across clusters. On top of that, the yield spread variable is significant only in the case of the last three clusters (that is, South-Asian countries), which indicates the importance of forward information to predict crises. These

results provide support for second generation models, as they stress the importance of non fundamental factors in the occurrence of financial turmoil.

From a graphical¹³ point of view, for Brazil, Indonesia, Korea, Taiwan, and Thailand, the probabilities of crisis obtained using a logit model that includes market expectation variables are generally superior to those obtained from a simple logit model. In contrast, for Latin American countries (namely, Argentina, Mexico, and Venezuela), the crises probabilities issued from the two models are quite close. Finally, for Peru, Philippines and Uruguay, the comparison is not easy, as each model outperforms the others, depending on the period. In other words, there are countries for which the model with market expectation variables seems better, while there are countries for which it is difficult to identify any permanent differences in the forecast of crisis probabilities; for these countries it is hard to choose a single best model. Nevertheless, the simple logit model is never the most adequate model for an entire period.

Generally speaking, all currency crisis episodes are forecasted quite well by these two models. For example, Figures 3 and 4 report the probabilities of having at least one crisis in the following 24 months obtained from the logit model that includes market expectation variables. Moreover, the yellow rectangles represent the periods for which at least one crisis occurred in the following 24 months, the blue rectangles stand for the real crisis periods, whereas the horizontal line is the optimal cut-off for each country.

insert Figures 3 and 4

It seems that warning signals of the Asian crisis are already apparent in 1995 in some of the concerned countries, while the 2002 Argentinean depreciation, the 1999 Brazilian crisis, the 1999 Peruvian crisis, and the 2002 Venezuelan crisis are anticipated with at least one year in advance. On the contrary, even if the 1998-2000 crisis in Philippines, and the 1997-1998 crisis in Thailand seem to have been forecasted by the model in due time, similar crisis probabilities in real crisis as well as calm periods diminish the chances of a correct interpretation of the results. Thus, it is not clear when steps should be taken to prevent a currency crisis from occurring. As for the crises that take place at the beginning of the sample, the 1994-1995 Mexican crisis is correctly identified by the model, while in contrast, the 1996 Venezuelan crisis is not identified by the model (the crisis probabilities are lower than the cut-off). We must add, though, that some other depreciations which are identified at the beginning of the estimation sample by our models are not quantified as crises by the KLR modified pressure index (for example, January 1995 in Argentina, 1995 in Brazil, 1992-1994 in Malaysia, and so on). Nevertheless, the majority of currency crisis episodes identified by these models has already been recorded and analyzed by other studies (Abiad, 1993; Dabrowski, 2003).

13. These results are available upon request.

Last but not least, all variables with the exception of the growth of domestic credit over GDP are significant in at least one model. Moreover, the use of robust standard errors has a certain effect on the significance of the coefficients, since some variables are no more significant, whereas others become highly significant.

The results of the simple logit model are quite similar to those of the logit model that includes market expectation variables in terms of the sign and significance of the common coefficients. At the same time, the Markov models are characterized by a large variability in terms of the sign and significance of the coefficients of the independent variables¹⁴.

To analyze the performance of each model, the optimal cut-off C must be estimated for each country. As explained in section 3.1, we use two graphical tools, namely, one based on the sensitivity-specificity indicators, and one based on several accuracy and error measures. Moreover, to compare our results with the existing literature, we also indicate the cut-off obtained when using the NSR criteria.

insert Tables 4 and 5

Tables 4 and 5 report the cut-off obtained for each country. We put in bold those cut-offs that maximize both sensitivity and specificity. It is noticeable that most of the time the optimal cut-off is the one obtained by the accuracy measures. Moreover, the cut-off obtained by NSR (reported as the KLR cut-off) is always higher than those resulting from the other two methods. This actually leads to larger specificities but lower sensitivities. Hence, it reveals that existing studies are much more conservative than our current study. Concerning the estimated models, the panel logit leads to a cut-off that varies between 0.08 and 0.38, whereas for Markov models, it is always higher and ranges between 0.7 and 1. These results highlight that the use of a predetermined cut-off may induce some loss of information and disturbances in the measure of the forecasting performance of the models.

Most importantly, the sensitivity and specificity values seem higher for logit models than for Markov ones; in the first case more than 60% of cases are correctly identified (see Table 4), while in the second case, this rate is lower, going down to 0% in some cases (see Table 5). In fact, there are even two models (namely, a Markov model with *resg* switching and Markov model that includes market expectation variables and *resg* switching) that deliver a constant cut-off of 0.9995 and sensitivities (specificities) of 0% (100%) for all countries.

At the same time, the market expectation logit model leads to higher sensitivity and specificity values for the South-Asian countries than the simple logit model, while the simple logit has higher values for the Latin-American countries, emphasizing the idea that the currency crises of the Asian countries can be better predicted by using market expectation variables. It thus confirms our previous results regarding the forward-looking specificity of

14. These results are available upon request from the authors

the Asian crisis. However, in the case of Markov models, the role of market expectation variables is not so clearly supported by the results.

It seems that logit models, despite their autocorrelation problem, outperform Markov models. Similarly, market expectation variables have a clearer effect when introduced in logit models. In the next subsection, the proper statistical assessment criteria for the performance of each model are presented.

5.2 Evaluation Criteria

The following statistics assess the performance of a model based on the sensitivity-specificity measures (AUC, Kuiper score, Pietra index, and Bayesian error rate) as well as by comparing the forecasts with the realizations of the crisis variable C24 (using criteria such as QPS and LPS). Therefore, the higher the value of AUC is the better the model will be. A positive value of Kuiper's score signifies that the model generates more hits than false alarms, and so its predictive performance should increase. Similarly, a higher Pietra index, and a lower Bayesian error rate indicate a more stringent model, as do values of QPS and LPS closer to zero.

insert Table 6

Table 6, shows that the logit model exhibits correct predictive properties. Specifically, AUC is higher than 0.731 ; the Kuiper score is always positive ; Pietra index values are greater than 0.145 ; the Bayesian error rate is less than 0.272 ; the absolute value of QPS is less than 0.37 ; and the absolute value of LPS is less than 0.558. The results of the two logit models (with or without forward looking variables) are quite similar, with small differences from one country to another.

Concerning Markov models, the results are less satisfactory. First of all, the two models with switching *resg* are not very different from a random model in terms of performance¹⁵: the AUC criterion very frequently equals 0.5, while Kuiper's score and Pietra index are very small. In addition, the Bayesian error rate, QPS, and LPS are high. Second, the other three models (*i.e.*, without *resg* switching) have similar behavior (see Table 6), and clearly outperform the first two (that is, those models with *resg* switching). Nevertheless, we get the overwhelming feeling that Markov specifications are outperformed by panel logit in terms of forecasting performances. Still, it has to be confirmed by proper statistical tests. This will constitute the final part of our analysis.

15. These results are available upon request.

5.3 Comparison Tests

To determine the best model specification, the three tests presented in section 3.2 are implemented. The first test uses the DM statistic to compare the forecasts of two non-nested models, while the second test uses the Clark-West MSPE-adj statistic to compare the forecasts of two competitive nested models. Finally, the third test compares the area under the ROC curve.

First, the ROC and Clark-West tests are used to compare the simple logit without and with market expectation variables; these results are presented in Table 7. They corroborate the idea that forward looking variables matter for currency crisis prediction. In fact, this conclusion is not surprising as it matches the results obtained with the predicted probability plots.

insert Table 7

In the case of Markov models, the differences between models are not always clear. For example, for countries like Indonesia, Korea, Malaysia, and Mexico, the Markov model with market expectation variables outperforms the simple Markov model (see Table 20). Similarly, for countries like Brazil, Indonesia, Peru and Uruguay, this Markov model with forward looking variables is outperformed by the Markov model with forward-looking variables and *spread* switching.

insert Table 8

Finally, the last stage in comparing the predictive abilities of the models consists in comparing the logit model with market expectation variables to the Markov model with market expectation variables and *spread* switching. Results reported in Table 8 indicate that the two tests are always significant, leaving no doubt concerning the choice of the panel logit as the best model. Thus, we can clearly conclude that the best model is the logit model with market expectation variables. Moreover, a sensitivity analysis of the Logit model with market expectation variables (see Appendix 2) demonstrates the robustness of this result.

6 Conclusion

This paper proposes to tackle both problems by developing a new statistical framework to evaluate EWS. Drawing on the traditional credit-scoring measures (Lambert and Lipkovich, 2008), our study moves beyond the simple analysis of NSR criteria and proposes a measure of accuracy as well as a sensitivity and specificity analysis. It is then possible to determine the optimal threshold for each country by relying on the sensitivity-specificity plot. Similarly,

we adjust the most important credit-scoring criteria (including AUC, Kuiper Score, Pietra Index, and the Bayesian Error Rate) to evaluate the performance of the model specifications we have developed. Finally, nested and non-nested comparison tests are developed to identify the optimal specification.

This approach provides a unified framework to compare candidate EWS models. In the assessment of the predictive power of panel logit and Markov switching frameworks, this framework leads to several conclusions. First, the panel logit model outperforms the Markov based EWS. Second, the introduction of forward looking variables (namely, term spread) clearly improves the forecasting abilities of the EWS. It thus confirms the adequacy of the second generation crisis models in explaining the occurrence of crises. Third, the optimal EWS based on a logit model with market expectation variables predicts quite well most currency crises in the specified emerging markets. Specifically, it forecasts correctly at least 67.9% of crises and 61% of calm periods in each of the 12 countries). The forecasting performance of this model and its robustness to some sensitivity analysis provides a dominating position within EWS models for the logit model with market expectation variables.

Such a conclusion is of course conditional on our specification. Hence, our analysis calls for deeper empirical applications that rely on some hypotheses (*i.e.*, through the use of two-regime models, autoregressive models and volatility regime dependent models,..) and including extra forward-looking variables (*i.e.*, to address the direction of the market). Nevertheless, our evaluation procedure provides a unified framework to compare EWS models and thus indicates a clear path for future research.

Bibliography

- (1) Abiad, A., (2003). Early Warning systems: A Survey and a Regime Switching Approach, IMF Working Paper.
- (2) Arias, G., and Erlandsson, G., (2005). Improving Early Warning Systems with a Markov Switching Model - an Application to South-East Asian Crises, Lund University, Working Paper.
- (3) Basel Committee on Banking Supervision, (2005). Studies on the Validation of Internal Rating Systems, working paper no.14, Bank for International Settlements.
- (4) Berg, A., and Pattillo, C., (1999). Predicting Currency Crises: The Indicators Approach and an Alternative. *Journal of International Money and Finance* 18, 561-586.
- (5) Berg, J.B., Candelon, B, and Urbain, J.P., (2008). A Cautious Note on the Use of Panel Models to Predict Financial Crises, *Economics Letters* 101, issue 1, 80-83.
- (6) Berg, A., and Cooke, R., (2004). Autocorrelation Corrected Standard Errors in Panel Probits: an Application to Currency Crisis Prediction, IMF Working Paper.
- (7) Bussiere, M., and Fratzscher, M., (2006). Towards a New Early Warning System of Financial Crises, *Journal of International Money and Finance* 25, issue 6, 953-973.
- (8) Clark, T. E., McCracken, M. W., (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, issue 1, 85-110.
- (9) Clark, T. E., West, K. D., (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138, issue 1, 291-311.
- (10) Diebold, F. X., Lee, J.-H. and Weinbach, G., (1994). *Regime Switching with Time-Varying Transition Probabilities*, in C.Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*. Oxford University Press.
- (11) Diebold, F. X., Mariano, S., (1995). Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, American Statistical Association 13, issue 3, 253-63.
- (12) Estrella, A., and Hardouvelis, G. A., (1991). The Term Structure as a Predictor of Real Economic Activity, *The Journal of Finance*, 46, issue 2, 555-576.
- (13) Estrella, A., Mishkin, F.S., (1996). The Yield Curve as a Predictor of US Recessions, *Current Issues in Economics and Finance* 2, issue 7, 1 - 5.
- (14) Estrella, A., and Trubin, M.R., (2006). The Yield Curve as a Leading Indicator : Some Practical Issues, *Current Issues in Economics and Finance* 12, issue 5, 1 - 7.
- (15) Fratzscher, M., (2003). On Currency Crises and Contagion, *International Journal of Finance and Economics*, 8(2), 109-129.

- (16) Fuertes, A.-M., Kalotychou, E., (2007) Optimal Design of Early Warning Systems for Sovereign Debt Crises, *International Journal of Forecasting*, 23(1), 85-100.
- (17) Gould, W., Pitblado, J., Sribney, W., Stata Corporation, (2005). *Maximum Likelihood Estimation with Stata*, Stata Press.
- (18) Granger, C.W.J., and Pesaran, M.H., (2000). Economic and Statistical Measures of Forecast Accuracy, *Journal of Forecasting* 19, 537-560.
- (19) Hamilton, J. D., (1995). *Regime Switching Methods*, University of California, Prepared for: Palgrave Dictionary of Economics.
- (20) Hamilton, J. D., (1988). Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates. *Journal of Economic Dynamics and Control* 12, 385-423.
- (21) Harding, D., Pagan, A., (2006). The Econometric Analysis of Constructed Binary Time Series, Working Papers Series 963, The University of Melbourne.
- (22) Hoeffding, W., (1948) A Class of Statistics with Asymptotically Normal Distributions. *Annals of Statistics*, 19, 293-325.
- (23) Hsiao, C., (2003). *Analysis of Panel Data*. Cambridge University Press.
- (24) Im, K.S. & Pesaran, M. H., & Shin, Y., (2003). Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115(1), 53-74.
- (25) Jacobs, J., Kuper, G., and Lestano, (2003). Indicators of Financial Crises do Work! An Early Warning System for Six Asian Countries, CCSO Centre for Economic Research.
- (26) Kaminsky, G., Lizondo, S., Reinhart, C., (1998). Leading Indicators of Currency Crises, IMF Staff Papers 45, issue 1, 1-48.
- (27) Kapetanios, G., (2003). Determining the Poolability of Individual Series in Panel Datasets, Working Paper 499.
- (28) Kim, C. J., (1994), Dynamic Linear Models With Markov Switching, *Journal of Econometrics* 64, 1-22.
- (29) Kumar, M., Moorthy, U. and Perraudin, W., (2003). Predicting Emerging Market Currency Crashes, *Journal of Empirical Finance* 10, 427-454.
- (30) Lambert, J., and Lipkovich, I., (2008). A Macro for Getting more out of your ROC Curve, SAS Global forum, paper 231.
- (31) Lestano and Jacobs, J., (2004). A Comparison of Currency Crisis Dating Methods: East Asia 1970-2002, CCSO Working Papers 200412, CCSO Centre for Economic Research.
- (32) Maddala, G.S., and Wu, S., (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, special issue, 631-652.

- (33) Martinez Peria, M.S.(2002) "A Regime-Switching Approach to the Study of Speculative Attacks : A Focus on EMS Crises", in J.D. Hamilton and B. Raj (eds), *Advances in Markov-Switching Models. Applications in Business Cycle Research and Finance*, Physica-Verlag, Heidelberg, Germany, 2002.
- (34) Obstfeld, M., (1984). *The Logic of Currency Crises*. *Cahiers Economiques et Monetaires* 43,189-213.
- (35) Williams, R., (2004). *A Note on Robust Variance Estimation for Cluster-Correlated Data*, *Biometrics* 56, Issue 2, 645 - 646.

Appendix 1 : A Robust Estimator of the Variance of the Parameters

As previously mentioned, we use a sandwich estimator to compute robust estimators of the variance for logit models. Technically, we know that the matrix of variance of the estimators is asymptotically equal to the inverse of the hessian matrix: $V(\hat{\beta}) = -H(\hat{\beta})^{-1}$. Still, this is appropriate only if we employ the real Data Generating Process (DGP). For a more permissive method from this point of view, we define the variance vector as follows:

$$V(\hat{\beta}) = (-H(\hat{\beta})^{-1})V(g(\hat{\beta}))(-H(\hat{\beta})^{-1}), \quad (33)$$

where $H(\hat{\beta})^{-1}$ is the inverse of the hessian matrix, and $V(g(\hat{\beta}))$ is the variance of the gradient. Using the empirical variance estimator of the gradient we find that:

$$V(\hat{\beta}) = -T/(T - 1)H(\hat{\beta})^{-1}\left\{\sum_{t=1}^T g_t(\hat{\beta})g_t(\hat{\beta})^{-1}\right\}(-H(\hat{\beta})^{-1}), \quad (34)$$

which is a robust variance estimator for the time-series model.

The main advantage of this sandwich method is that it can also be applied in the case of grouped data, as in our case. It is important to note that in the current situation, each country from a cluster is a group of time-series observations that are correlated. Thus, the observations corresponding to a country are not treated as independent, but rather the countries themselves which form the clusters, are considered independent. Therefore, instead of using $g_t(\hat{\beta})$, we use the sum of $g_t(\hat{\beta})$ for each country, while T is replaced by the number of countries in a cluster. These changes ensure the independence of so-called "superobservations" entering the formula (Gould et al., 2005).

Appendix 2 : A Robustness Check of the Optimal Model

In this section, we propose a sensitivity analysis of the performance of the best EWS for currency crises with respect to the dating of currency crises as well as to the forecasting horizon.

First, instead of using the KLR modified pressure index (KLRm), we will use the Zhang original index (ZCC) (Lestano and Jacobs, 2004). It is defined as follows:

$$Crisis_{n,t} = \begin{cases} 1, & \text{if } \begin{cases} \frac{\Delta e_{n,t}}{e_{n,t}} > \beta_1 \sigma'_{e_{n,t}} + \mu_{e_{n,t}} & \text{or} \\ \frac{\Delta r_{n,t}}{r_{n,t}} < \beta_2 \sigma'_{r_{n,t}} + \mu_{r_{n,t}} \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (35)$$

where $\sigma'_{e,t}$ is the standard deviation of $(\Delta e_{n,t}/e_{n,t})$ in the sample of (t-36, t-1), and $\sigma'_{r,t}$ is the standard deviation of $(\Delta r_{n,t}/r_{n,t})$ in the sample of (t-36, t-1). The thresholds are set to $\beta_1 = 3$ and $\beta_2 = -3$. Contrary to the KLRm index, the interest rates are excluded from the ZCC and the thresholds used are time-varying for each component.

Second, the crisis variable is defined using C_{12} instead of C_{24} ; *i.e.*, it takes the value of 1 if there is a crisis in the following 12 months, and the value of 0 otherwise.

It appears that both new estimations of the model resemble the estimators generated from the original logit model with market expectation variables in terms of global goodness of fit, and the sign and significance of the parameters. Still, some differences appear, especially when considering the ZCC index. Significant variables are less numerous and are different from the original model, including M2 to reserves and growth of stock market price index; there is also a more intense variability of parameter signs from one cluster to another. In contrast, the results obtained using C_{12} are strictly identical to the ones derived from the initial logit model with market expectation variables.

The optimal cut-off for each country is reported in Table 9.

insert Table 9

It appears that the optimal cut-off is almost always given by the accuracy criteria. In addition, it is always inferior to 0.35 (except for Uruguay). The sensitivity and specificity values of the three models are quite similar, and they are particularly good for countries like Mexico, Peru (ZCC) as well as Argentina, Korea, Malaysia, Peru, Philippines, Thailand, Uruguay, Venezuela (C_{12}). This conclusion is supported by the performance assessment criteria (see Tables 10 and 11).

insert Table 10 and 11

All in all, this robustness check supports our main findings.

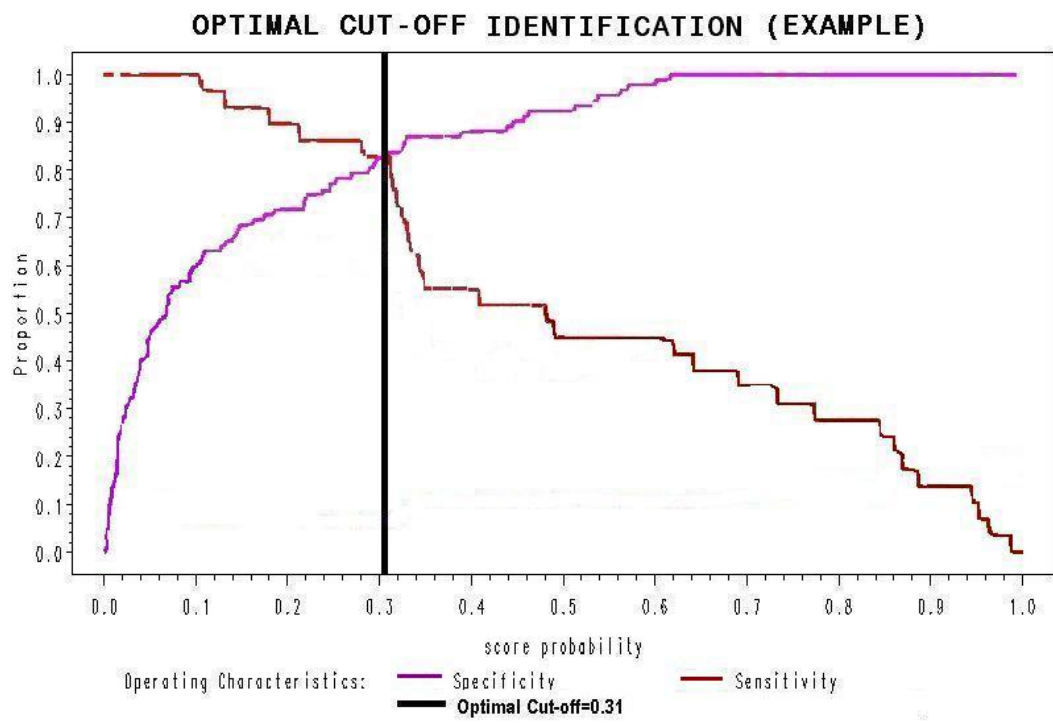


FIGURE 1 – Optimal Cut-off determination

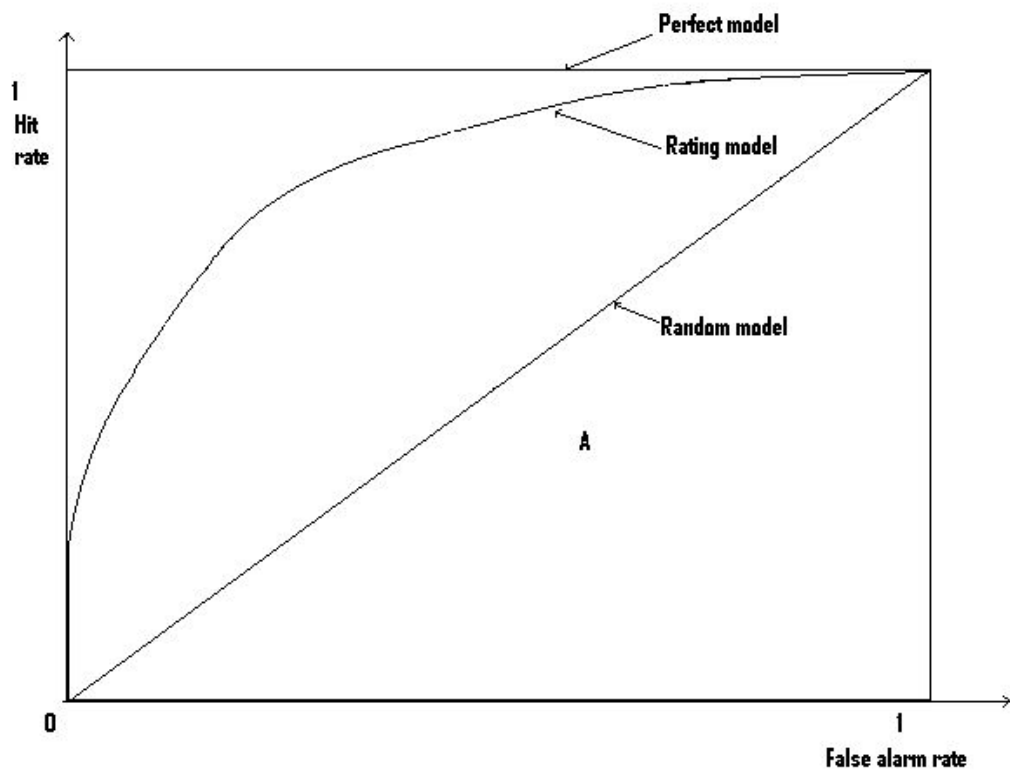
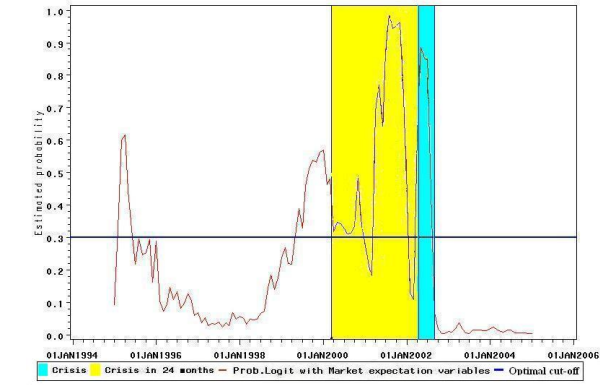
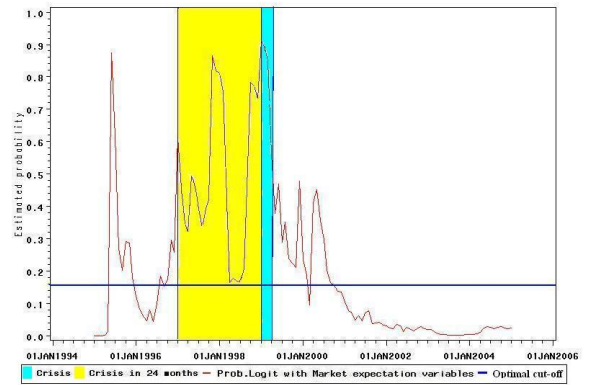


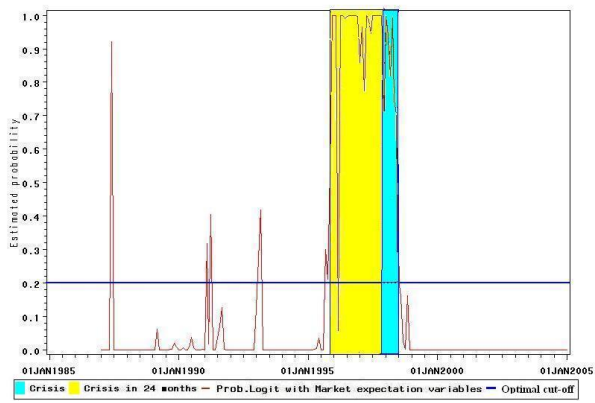
FIGURE 2 – The ROC curve



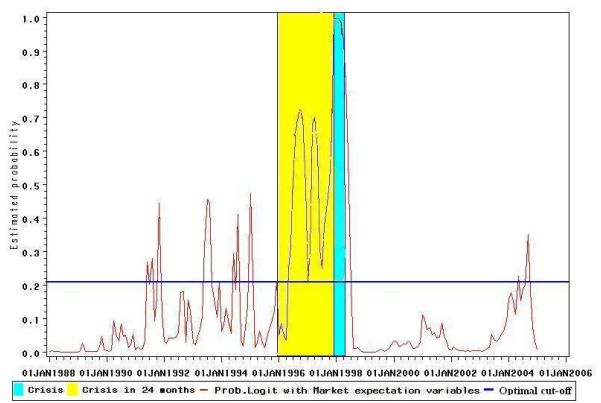
(a) Argentina



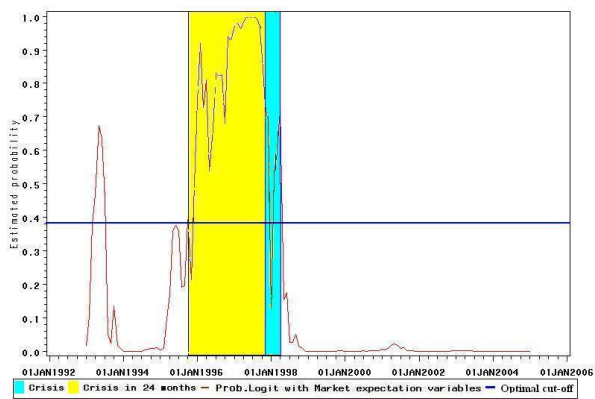
(b) Brazil



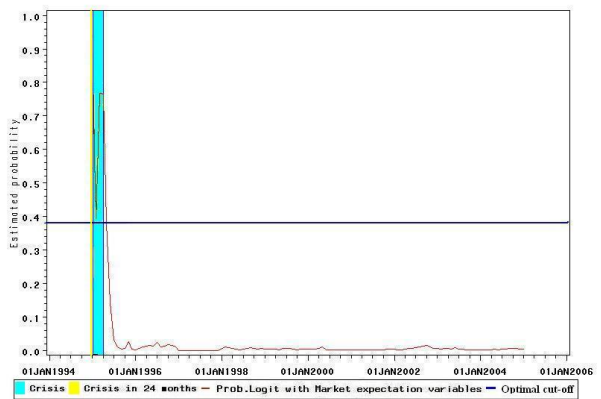
(c) Indonesia



(d) Korea

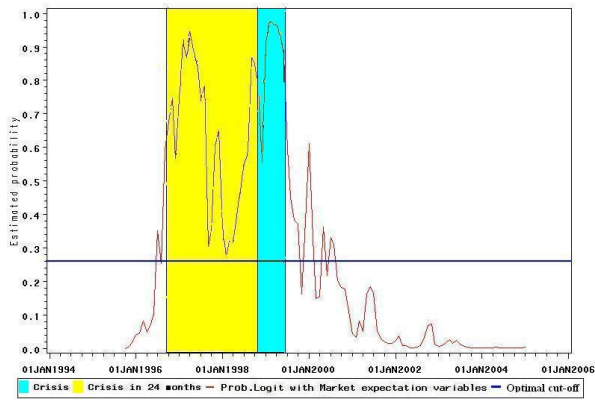


(e) Malaysia

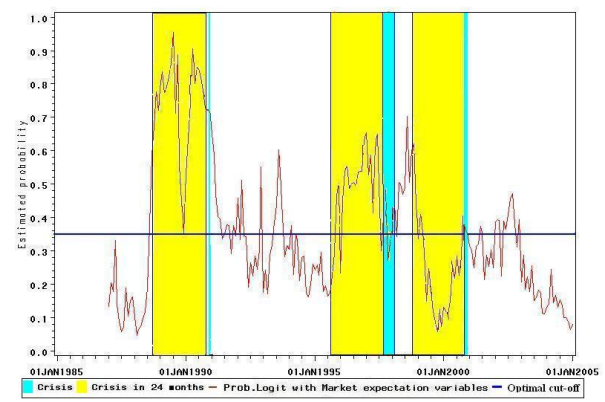


(f) Mexico

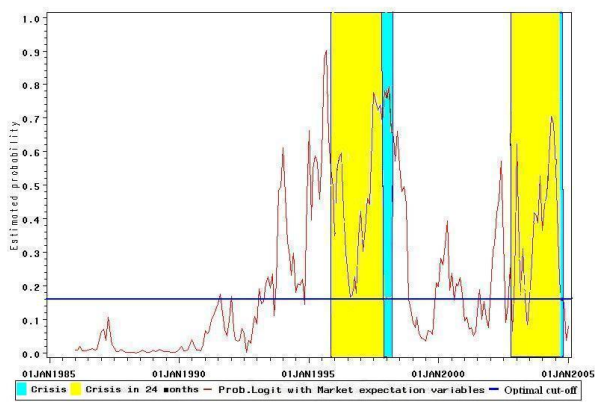
FIGURE 3 – Predicted probability of crisis



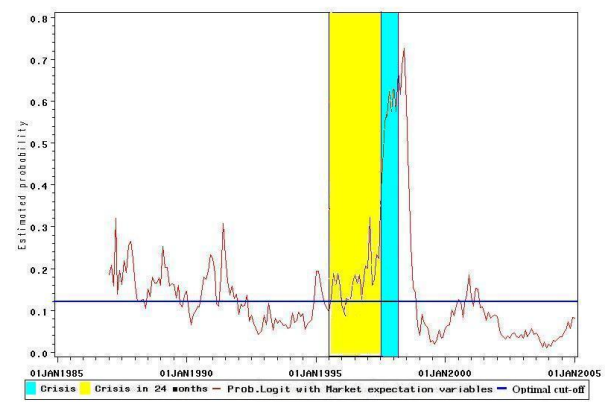
(a) Peru



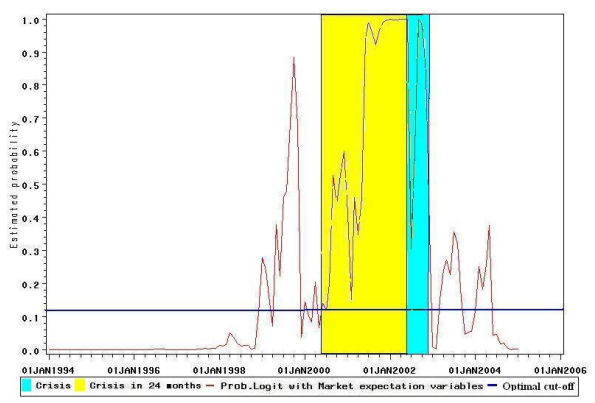
(b) Philippines



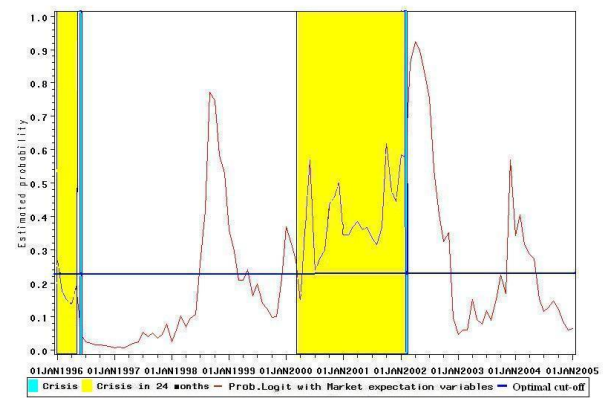
(c) Taiwan



(d) Thailand



(e) Uruguay



(f) Venezuela

FIGURE 4 – Predicted probability of crisis (continued)

TABLE 2 – Estimation results of a logit model with market expectation variables

Indicator	cluster 1		cluster 2		cluster 3	
	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Growth of international reserves	-1.426	2.374	-0.817	3.452	-9.190	1.327
Growth of mm2 multiplier	-1.136	4.411	8.192	0.359	-3.678	4.565
Growth of domestic credit over GDP	-2.787	3.770	2.594	5.469	4.756	6.983
Real interest rate	18.72	6.125	41.04	21.37	-4.248	17.56
First difference of lending rate over deposit rate	0.631	0.284	-0.374	0.660	-0.868	0.526
Growth of real bank deposits	-2.809	0.822	-2.699	5.698	4.572	0.260
M2 to reserves	3.454	1.704	-20.35	4.448	25.25	4.002
Growth of M2 to reserves	-7.39	0.562	-10.02	1.796	-11.16	1.369
Growth of industrial production	1.769	1.218	-1.813	5.903	5.198	4.655
Growth of stock market price index	0.549	1.068	2.755	0.403	0.376	1.240
Yield Spread	0.299	0.358	-0.234	0.439	-1.545	0.343
Likelihood Ratio	140.644		147.501		234.707	
Log pseudolikelihood	-129.995		-58.3494		-150.614	
Pseudo R2	0.35110		0.55830		0.43790	
AIC	281.991		138.699		323.228	
SC	327.718		177.213		371.202	

TABLE 3 – Estimation results of a logit model with market expectation variables (continued)

Indicator	cluster 4		cluster 5	
	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Growth of international reserves	-0.287	2.308	-12.23	14.06
Growth of mm2 multiplier	1.822	8.984	23.50	11.82
Growth of domestic credit over GDP	0.527	2.389	7.981	5.727
Real interest rate	-1.731	6.259	11.38	20.06
First difference of lending rate over deposit rate	-0.592	0.465	4.410	10.50
Growth of real bank deposits	0.165	1.073	38.47	10.63
M2 to reserves	4882	354.0	$6.14 * 10^6$	$4.6 * 10^6$
Growth of M2 to reserves	-2.107	3.149	-40.04	31.90
Growth of industrial production	-4.242	0.336	-41.16	21.14
Growth of stock market price index	-1.155	2.388	4.716	4.308
Yield Spread	-0.221	0.018	-4.157	1.614
Likelihood Ratio	67.5548		159.626	
Log pseudolikelihood	-191.503		-12.6918	
Pseudo R2	0.14990		0.86280	
AIC	405.006		49.3840	
SC	449.809		89.9420	

TABLE 4 – Optimal cut-off identification in a Logit model with market expectation variables

Country	Accuracy measures			Sensitivity-Specificity graphic			KLR		
	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
Argentina	0.300	82.76	82.61	0.300	82.76	82.61	0.620	41.38	100.0
Brazil	0.160	100.0	69.47	0.250	80.77	81.05	0.880	7.69	100.0
Indonesia	0.200	96.97	96.20	0.210	96.97	96.20	0.930	72.73	100.0
Korea	0.206	85.71	90.96	0.170	85.71	85.31	0.930	14.29	100.0
Malaysia	0.380	93.10	93.97	0.370	93.10	93.10	0.730	65.52	100.0
Mexico	0.379	100.0	99.15	0.379	100.0	99.15	0.390	75.00	100.0
Peru	0.260	100.0	82.72	0.350	87.10	86.42	0.940	12.90	100.0
Philippines	0.400	64.10	82.01	0.346	67.95	68.35	0.730	20.51	100.0
Taiwan	0.160	94.12	65.17	0.228	76.47	76.97	0.670	17.65	98.31
Thailand	0.120	90.32	61.29	0.150	74.19	74.19	0.321	25.81	96.24
Uruguay	0.119	93.33	75.73	0.230	83.33	83.50	0.900	50.00	100.0
Venezuela	0.225	85.71	67.90	0.280	71.43	71.60	0.330	64.29	77.78

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold. For comparison reasons, we also present the KLR cut-off which is obtained by minimizing NSR (reported as KLR).

TABLE 5 – Optimal cut-off identification in a Markov model with market expectation variables and *spread* switching

Country	Accuracy measures			Sensitivity-Specificity graphic			KLR		
	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
Argentina	0.990	51.72	36.96	0.990	51.72	36.96	0.987	100.0	<0.001
Brazil	0.800	65.38	73.68	0.610	65.38	65.26	0.964	50.00	85.26
Indonesia	0.522	57.58	56.52	0.522	57.58	56.52	0.828	24.24	98.37
Korea	0.500	71.43	71.19	0.500	71.43	71.19	0.947	39.29	87.57
Malaysia	0.967	100.0	58.62	0.988	68.97	71.55	0.989	44.83	86.21
Mexico	0.923	50.00	51.28	0.923	50.00	51.28	0.922	75.00	47.86
Peru	0.860	41.94	59.26	0.85	100.0	<0.001	0.982	32.26	75.31
Philippines	0.928	60.26	66.19	0.910	61.54	61.87	0.926	60.26	66.19
Taiwan	0.952	64.71	42.70	0.980	50.98	50.56	0.953	64.71	43.26
Thailand	0.875	74.19	23.66	0.890	38.71	34.41	0.989	19.35	91.40
Uruguay	0.520	100.0	<0.001	0.527	33.33	26.21	0.524	100.0	<0.001
Venezuela	0.988	82.14	29.63	0.989	17.86	58.02	0.988	82.14	29.63

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold. For comparison reasons, we also present the KLR cut-off which is obtained by minimizing the noise-to-signal ratio (reported as KLR).

TABLE 6 – Evaluation criteria for a logit model with market expectation variables and a Markov model with market expectation variables and *spread* switching

Country	Model	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	Logit	0.898	65.37	0.235	0.132	0.215	-0.325
	Markov	0.625	-11.32	<0.001	0.239	1.492	3.558
Brazil	Logit	0.907	69.47	0.249	0.132	0.202	-0.311
	Markov	0.710	39.06	0.142	0.215	0.765	1.057
Indonesia	Logit	0.996	93.17	0.330	0.0138	0.034	-0.058
	Markov	0.685	14.10	0.114	0.129	0.567	0.607
Korea	Logit	0.920	76.67	0.273	0.0780	0.135	-0.228
	Markov	0.748	42.62	0.164	0.136	0.682	0.909
Malaysia	Logit	0.985	87.07	0.311	0.048	0.083	-0.131
	Markov	0.809	58.62	0.207	0.200	1.184	2.165
Mexico	Logit	0.998	99.15	0.350	0.008	0.011	-0.023
	Markov	0.564	1.280	0.081	0.033	1.658	2.526
Peru	Logit	0.947	82.72	0.292	0.107	0.166	-0.266
	Markov	0.529	<0.001	0.027	0.277	1.192	1.939
Philippines	Logit	0.739	36.30	0.163	0.235	0.368	-0.558
	Markov	0.582	23.41	0.093	0.359	1.080	1.756
Taiwan	Logit	0.837	59.29	0.211	0.196	0.270	-0.399
	Markov	0.511	7.410	0.028	0.223	1.467	3.085
Thailand	Logit	0.811	51.61	0.192	0.138	0.218	-0.348
	Markov	0.592	-2.150	0.038	0.143	1.511	2.710
Uruguay	Logit	0.939	69.06	0.257	0.105	0.165	-0.246
	Markov	0.725	<0.001	<0.001	0.225	1.186	2.023
Venezuela	Logit	0.777	53.61	0.189	0.257	0.370	-0.530
	Markov	0.511	11.77	0.042	0.257	1.454	3.361

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞ , 0 being perfect accuracy.

TABLE 7 – Comparison tests of Simple logit and Market expectation logit models

Country	ROC		Clark-West	
	test statistic	p-value	test statistic	pvalue
Argentina	0.0301	0.8622	0.1372	0.4454
Brazil	5.7105	0.0169	3.4901	0.0002
Indonesia	7.9917	0.0047	4.4332	0.0000
Korea	4.5357	0.0332	3.7746	0.0001
Malaysia	0.3859	0.5345	0.3288	0.3711
Mexico	<0.001	1.0000	0.6869	0.2460
Peru	0.0028	0.9577	2.1634	0.0153
Philippines	0.8738	0.3499	0.8709	0.1919
Taiwan	10.475	0.0012	3.5603	0.0002
Thailand	6.9801	0.0082	4.5964	0.0000
Uruguay	0.7443	0.3883	0.6656	0.2528
Venezuela	6.6647	0.0098	-2.0740	0.9810

Note : The null hypothesis of the ROC test is the equality of areas under the ROC curve, while the alternative hypothesis is the statistical difference between the two areas. Its statistic follows a normal distribution whose critical values are ± 1.96 (5%). The null hypothesis of the Clark-West test is the equality of predictive performance of the two models. The alternative indicates that the non-constraint model (the bigger one) is better than the other one. Under the null hypothesis, the MSPE-adj statistic follows a normal distribution with a critical unilateral value of 1.645(5%). Bold entries indicate significance at the 5% level.

TABLE 8 – Comparison tests of market expectation logit and market expectation *spread* switching Markov models

Country	ROC		Diebold-Mariano	
	test statistic	p-value	test statistic	pvalue
Argentina	62.678	<0.001	12.965	<0.001
Brazil	9.7859	0.0018	8.783	<0.001
Indonesia	46.529	<0.001	29.244	<0.001
Korea	9.8754	0.0017	12.207	<0.001
Malaysia	21.455	<0.001	17.066	<0.001
Mexico	17.829	<0.001	50.850	<0.001
Peru	45.942	<0.001	12.164	<0.001
Philippines	7.4266	0.0064	9.7129	<0.001
Taiwan	34.195	<0.001	16.591	<0.001
Thailand	45.902	<0.001	18.281	<0.001
Uruguay	125.00	<0.001	12.877	<0.001
Venezuela	17.351	<0.001	9.4665	<0.001

Note : The null hypothesis of the ROC test is the equality of area under the ROC curve, while the alternative hypothesis is the statistical difference between the two areas. Its statistic follows a normal distribution whose critical values are ± 1.96 (5%). The null hypothesis of the Diebold-Mariano test is the equality of predictive performance of the two models. The alternative indicates that the first model is better than the other one. Under the null hypothesis, the test statistic follows a normal distribution. Bold entries indicate significance at the 5% level.

TABLE 9 – Optimal cut-off identification in a logit model with market expectation variables

Country	Zhang method						C12 method					
	Accuracy measuress			Sensitivity-Specificity graphic			Accuracy measuress			Sensitivity-Specificity graphic		
	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity	Cut-off	Sensitivity	Specificity
Argentina	0.080	97.14	58.00	0.550	65.71	70.00	0.180	94.12	90.38	0.195	88.24	90.38
Brazil	0.330	100.0	68.09	0.438	76.32	78.72	0.060	92.86	61.68	0.080	71.43	71.03
Indonesia	0.200	100.0	100.0	0.200	100.0	1000	0.3500	100.0	100.0	0.3500	100.0	100.0
Korea	0.150	85.19	78.87	0.170	81.48	81.69	0.110	93.75	89.95	0.150	93.75	93.65
Malaysia	0.030	100.0	96.47	0.038	95.83	96.47	0.080	100.0	90.63	0.120	94.12	93.75
Mexico	0.320	79.17	78.69	0.340	79.17	80.33	0.225	100.0	98.29	0.225	100.0	98.29
Peru	0.020	75.00	73.61	0.020	75.00	73.61	0.110	94.74	91.40	0.260	94.74	94.62
Philippines	0.220	82.05	85.21	0.200	82.05	82.39	0.180	71.43	70.29	0.180	71.43	70.29
Taiwan	0.200	100.0	95.81	0.220	96.15	95.81	0.110	85.19	80.69	0.120	81.48	81.68
Thailand	0.200	100.0	64.8	0.350	78.57	78.4	0.120	94.74	83.84	0.160	89.47	88.89
Uruguay	0.685	26.79	95.2	0.700	91.53	89.47	0.300	94.44	97.39	0.120	94.44	93.91
Venezuela	0.240	92.86	68.8	0.450	79.31	81.82	0.060	100.0	58.06	0.140	75.00	75.27

Note : For each country we choose the optimal cut-off from the values obtained by using two different methods (accuracy measures and sensitivity-specificity graphic) as being the one that maximizes both sensitivity and specificity, usually giving more weight to the correct identification of crisis periods (sensitivity). The selected cut-off values are in bold.

TABLE 10 – Evaluation criteria for a logit model with market expectation variables (ZCC)

Country	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	0.797	55.14	0.195	0.259	0.389	-0.548
Brazil	0.895	68.09	0.241	0.176	0.274	-0.419
Indonesia	1.000	100.0	0.353	<0.001	0.001	-0.005
Korea	0.887	64.06	0.229	0.094	0.169	-0.282
Malaysia	0.999	96.47	0.341	0.009	0.025	-0.042
Mexico	0.878	59.5	0.213	0.188	0.279	-0.432
Peru	0.892	48.61	0.245	0.052	0.090	-0.154
Philippines	0.914	67.26	0.240	0.105	0.199	-0.339
Taiwan	0.994	95.81	0.339	0.031	0.046	-0.075
Thailand	0.88	64.80	0.229	0.177	0.267	-0.405
Uruguay	0.968	81.00	0.298	0.072	0.129	-0.214
Venezuela	0.897	61.66	0.257	0.151	0.256	-0.395

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞ , 0 being perfect accuracy.

TABLE 11 – Evaluation criteria for a logit model with market expectation variables (C12)

Country	AUC	Kuiper score	Pietra index	Bayesian error rate	QPS	LPS
Argentina	0.956	84.50	0.299	0.049	0.095	-0.171
Brazil	0.861	54.54	0.203	0.083	0.153	-0.257
Indonesia	1.000	100.0	0.354	<0.001	0.002	-0.005
Korea	0.973	83.70	0.316	0.044	0.073	-0.113
Malaysia	0.992	90.63	0.327	0.021	0.045	-0.077
Mexico	0.996	98.29	0.347	0.008	0.017	-0.028
Peru	0.987	89.36	0.323	0.036	0.068	-0.115
Philippines	0.755	41.72	0.149	0.152	0.264	-0.425
Taiwan	0.898	65.88	0.233	0.087	0.142	-0.234
Thailand	0.945	78.58	0.279	0.059	0.105	-0.177
Uruguay	0.989	91.83	0.325	0.023	0.045	-0.092
Venezuela	0.823	58.06	0.205	0.138	0.224	-0.338

Note : The AUC criteria takes values between 0.5 and 1, 1 being the perfect model. Kuiper's score should have positive values if the model identifies well the crisis periods. Pietra index takes values from -0.354 to 0.354, the higher its level, the better the model. Bayesian error rate takes values between 0 and 1, 0 corresponding to the perfect model. QPS ranges from 0 to 2, 0 being perfect accuracy, while LPS ranges from 0 to ∞ , 0 being perfect accuracy.